



María Teresa Blaconá*
María del Carmen García*
María Gabriela Borgognone*
Javier Bussi*
Nora Ventroni*
José Luis Pellegrini**

**Instituto de Investigaciones Teóricas y Aplicadas de la Escuela de Estadística. Escuela de Estadística.*

*** Instituto de Investigaciones Económicas. Escuela de Economía.*

CONSIDERACIONES METODOLÓGICAS SOBRE LA ESTIMACIÓN ECONOMÉTRICA DE LAS ECUACIONES DE INGRESOS DE LOS INTEGRANTES DE LA PAREJA CONYUGAL

1. Introducción

La conocida ecuación de ingresos del capital humano de Mincer (1970) ha sido la base para la especificación de modelos econométricos en numerosos estudios realizados en todo el mundo sobre rendimientos de la educación, retornos del entrenamiento laboral y diferencias salariales, entre otros.

Aún hoy existen frecuentes discusiones sobre los factores que intervienen en la formación del ingreso y los métodos con que se los estima. Sin embargo, en general, los trabajos econométricos recientes sobre estos temas y otros análogos realizados en nuestro país, también han adoptado la teoría del capital humano (Becker, 1983), a veces con un enfoque un tanto ingenuo. No obstante, generalmente se obtienen estimaciones que proporcionan los signos esperados para las relaciones entre las variables relevantes.

Una vertiente teóricamente afín de la investigación empírica, mucho menos explorada en la Argentina, se ha centrado en la explicación económica de los comportamientos familiares. En la tradición de Becker (1973, 1974), una de las hipótesis centrales de la teoría económica del matrimonio es la de la formación de parejas basada en la asociación por características personales¹, entre las que se destacan los ingresos y sus principales determinantes según la teoría del capital humano. Asimismo, se observó que los salarios de los maridos se asocian positivamente con la educación de la esposa.

Algunas de las diferentes explicaciones que se propusieron para este fenómeno, postulan que la educación de la esposa (Beham citado por Becker, 1983) o las horas de trabajo de ésta (Jacobsen and Rayack, 1996) afectan la productividad de mercado del marido, por lo que el ingreso del varón no sería independiente de las características de la esposa. Por otra parte, si se admite que los integrantes del hogar comparten el mismo presupuesto, el comportamiento de ambos esposos en el mercado laboral debería estar relacionado.

En los estudios econométricos, cuando se consideran algunas de las características del cónyuge en una ecuación de ingresos individual, muchas veces se las introduce como variables exógenas, al igual que la propia educación y otras variables de capital humano. Ese criterio fue objetado tempranamente por Griliches (1977), en relación con la educación, y por Boulier y Rosenzweig (1984), en lo que se refiere a las características llevadas por las personas al mercado matrimonial y tratadas como si fueran exógenas. Lam (1988) ha

¹ Se prefirió utilizar esta expresión como traducción al castellano del término *assortative mating*, porque aunque menos sintética, parece más descriptiva que "emparejamiento selectivo", utilizada en la difundida versión española del *Treatise on the Family* publicada por Alianza Editorial (Becker, 1983).

resaltado que una implicación empírica de cualquier modelo de ordenamiento marital es que las características de los esposos son endógenas al mercado matrimonial. Esto debería tenerse en cuenta a la hora de estimar las ecuaciones de ingreso individual.

Otra limitación que suele encontrarse en los estudios de corte transversal se relaciona con la posible influencia de variables inobservables. Behrman et al. (1995), en un estudio de datos de panel, encontraron evidencia de la existencia de características no observadas por el investigador, pero sí por los intervinientes, que son valiosas tanto en el mercado matrimonial como en el laboral, y que permiten a los maridos elegir esposas más educadas y con mejores salarios. La consideración conjunta de los ingresos de los dos integrantes de la pareja conyugal, mediante la estimación de un sistema de ecuaciones de ingreso simultáneas, puede ayudar a atenuar el problema de endogeneidad señalado, al tiempo que permite tener en cuenta las posibles relaciones entre los ingresos de ambos.

El signo de la asociación de los salarios entre ambos cónyuges ha estado sujeto a controversia. Los salarios en las parejas de dos perceptores están correlacionados, incluso cuando se controla por educación, edad, experiencia, etc. (Lam, 1988). Sin embargo, Becker (1974) había sostenido que esa evidencia no constituía una buena prueba de su predicción teórica de que el signo de la asociación era negativo, debido a que es más probable que las mujeres con salarios más elevados, y en consecuencia más próximos a los de su marido, sean económicamente activas, a diferencia a aquéllas de salarios más bajos, por lo que éstos son inobservables. Por este motivo, los problemas de censura deben ser tratados adecuadamente, ya que al no considerarlos pueden obtenerse resultados expuestos a crítica.

En este trabajo se estiman, en forma separada y conjunta, ecuaciones de ingresos laborales individuales de los integrantes de la pareja conyugal, empleando diferentes modelos y métodos econométricos. El propósito es plantear una discusión introductoria sobre las dificultades que se presentan por la posible endogeneidad de algunas variables explicativas y la censura de los datos. Se comparan diferentes métodos utilizados para la estimación de modelos que explican el ingreso de cada uno de los integrantes de la pareja conyugal, planteando un modelo estructural para la Argentina, a partir de la teoría económica existente sobre los ingresos individuales. La aplicación se realiza a partir de la información sumistrada por la Encuesta Permanente de Hogares (EPH), relevada por el INDEC para el Conurbano Bonaerense, correspondiente a la primera onda de 1998.

2. Algunas consideraciones al estimar modelos econométricos de ingreso

Existen numerosos trabajos en la literatura nacional e internacional sobre modelos que intentan explicar la formación del ingreso individual. Estos trabajos tratan problemas tanto teóricos como prácticos del proceso de identificación y estimación de los mismos. Publicaciones de los últimos años advierten sobre los peligros de omitir características individuales que estén correlacionadas con los salarios porque pueden sesgar las estimaciones que se obtienen de secciones transversales. Tal es el caso, por ejemplo, para las estimaciones obtenidas por mínimos cuadrados ordinarios (Berlinski, 2000).

Las estimaciones realizadas en el país mediante el análisis de regresión han proporcionado generalmente coeficientes significativos y con los signos esperados en virtud de la Teoría del Capital Humano, excepto en el caso de la "experiencia" (medida según Mincer), para la cual resultan, a menudo, no significativos. Otro de los aspectos que generalmente no se ha tenido en cuenta en los estudios realizados en la Argentina ha sido el de la relación entre los logros educativos de uno de los miembros de la pareja conyugal y los ingresos del otro miembro.

Becker (1973) propuso una aplicación fundada en la teoría de la formación de parejas basada en la asociación de características. Sugirió que la educación de la esposa es una

variable *proxy* de características no observadas que afectan la productividad del cónyuge en las actividades fuera del mercado.

La existencia de una relación positiva entre ambas variables había sido previamente señalada por Benham (citado por Becker, 1983), en estudios realizados en Estados Unidos. Sin embargo, Jacobsen y Rayack (1996) no encontraron diferencias substanciales entre los ingresos de los maridos independientemente del hecho que sus esposas trabajaran o no. Si la educación fuera una característica valiosa en el mercado matrimonial, la introducción de las características de un cónyuge en las ecuaciones del ingreso del otro cónyuge permitiría, eventualmente, reducir el número de variables omitidas (e incorporar una variable *proxy* de sus habilidades).

Mouchart et al. (1999) desarrollaron un modelo con una única ecuación para datos de panel con información sobre el GBA extraída de la EPH para las ondas desde mayo de 1996 hasta octubre de 1997 y encontraron que la educación está aparentemente relacionada con los residuos. Ésto, de acuerdo a los autores, proporciona estimaciones inconsistentes y sugieren incorporar la variable educación como endógena y trabajar con un sistema de ecuaciones.

Griliches (1977), refiriéndose al problema de la correlación de la educación con los residuos, señala que la escolarización es el resultado, al menos en parte, del comportamiento de optimización de los individuos y de sus familias. Este comportamiento se basa en alguna función de ingreso anticipado. Como consecuencia, los errores de las funciones de ingreso ex-post y ex-ante están correlacionados. Ello afectaría también a la ecuación de escolarización e induciría a una correlación adicional entre esta variable y los errores, lo que avalaría el uso de métodos de ecuaciones simultáneas estimando el coeficiente de escolarización en tal sistema.

De lo expuesto surge que el mecanismo socioeconómico de la formación de ingresos individuales es complejo. Para Argentina no existen hasta el momento estudios sobre el tema que tengan en cuenta el carácter dinámico del fenómeno, las posibles diversidades existentes entre las distintas regiones del país y la endogeneidad de algunas variables que lo explican.

3. Metodología

En muchos casos interesa explicar el comportamiento simultáneo de ciertas variables mediante la formulación de un conjunto de ecuaciones que reflejen las relaciones entre las mismas. Los modelos de ecuaciones simultáneas, utilizados para tal fin, se caracterizan por la presencia de variables endógenas (respuesta) entre las variables explicativas de algunas o de todas las ecuaciones y la presencia de correlación de algunos regresores con los errores de la ecuación de regresión. Este último efecto, denominado endogeneidad de los regresores, puede causar serios problemas al realizar inferencias en econometría. Los estimadores de los parámetros, además, resultan inconsistentes al utilizar mínimos cuadrados ordinarios (MCO), por lo tanto se requieren otros métodos de estimación con variables instrumentales, como mínimos cuadrados a dos o tres etapas.

Un problema que surge a menudo es la presencia de datos censurados. En estos casos no es conveniente aplicar los métodos de estimación corrientes, pues los estimadores resultan ineficientes. Un método alternativo es el método de regresión censurada propuesto por Tobin (1958), llamado generalmente Tobit.

3.1. El modelo de regresión censurada: Análisis Tobit.

Al trabajar con datos microeconómicos es frecuente que la variable de interés tenga problemas de censura. Cuando la respuesta está censurada, todos los valores contenidos

en un cierto rango se transforman en un único valor (o se presentan como si fueran iguales a ese valor, el cual puede ser cero). Tal es el caso, por ejemplo, del ingreso individual de la mujer, horas trabajadas, etc.

Cuando la variable respuesta toma el valor cero en un gran porcentaje de las observaciones no se puede explicar, con los métodos de regresión clásicos, la diferencia cualitativa entre las observaciones *límite* (cero) y las observaciones *no límite* (continuas). Si existen valores censurados, la distribución que siguen los datos de la muestra es una mezcla entre una distribución continua y otra discreta.

Para encontrar la distribución de una variable censurada se define una nueva variable aleatoria y^* , obtenida a partir de la original de la siguiente manera:

$$\begin{aligned} y^* &= 0 \quad \text{si } y \leq 0, \\ y^* &= y \quad \text{si } y > 0. \end{aligned} \quad (3.1.1)$$

Si se considera que $y \sim N(\mu, \sigma^2)$, la distribución de y^* viene dada por

$$\begin{cases} \text{Prob}(y^* = 0) = \text{Prob}(y \leq 0) = \Phi\left(-\frac{\mu}{\sigma}\right) = 1 - \Phi\left(\frac{\mu}{\sigma}\right) & y \\ f(y^*) = f(y) & \text{si } y > 0 \end{cases} \quad (3.1.2)$$

El modelo de regresión basado en una variable que presenta censura se denomina modelo de regresión censurada o modelo Tobit. La formulación general suele darse en términos de una función índice, tomando a la media de la variable censurada como equivalente a un modelo de regresión clásico. El modelo queda expresado como:

$$\begin{aligned} y_i &= x_i' \beta + \epsilon_i \\ y_i^* &= 0 \quad \text{si } y_i \leq 0 \\ y_i^* &= y_i \quad \text{si } y_i > 0 \end{aligned} \quad (3.1.3)$$

En este modelo hay, en principio, tres funciones de media condicional y el objetivo del estudio determina cuál de ellas analizar.

1. La función media condicional de la variable índice, denominada a veces variable latente, es $E(y_i) = x_i' \beta$. Sin embargo, si los datos siempre presentan censura es probable que esta función no tenga ninguna utilidad.

2. La esperanza de una observación extraída aleatoriamente de la población, que puede estar censurada o no, es

$$E(y_i^* / x_i) = \Phi\left(\frac{x_i' \beta}{\sigma}\right)(x_i' \beta + \sigma \lambda_i), \quad \text{siendo } \lambda_i = \frac{\phi(x_i' \beta / \sigma)}{\Phi(x_i' \beta / \sigma)} \quad (3.1.4)$$

3. Si la intención es centrarse en las observaciones no censuradas, cuya esperanza condicional viene dada por $E(y_i^* / x_i) = x_i' \beta + \sigma \lambda_i$, se utilizan los métodos desarrollados para variables truncadas.

No hay un acuerdo en la literatura sobre qué función de media condicional usar para predecir observaciones futuras, pero la lógica sugeriría la ecuación (3.1.4).

Los efectos marginales también difieren si se utiliza una u otra función de media condicional.

Así, si se usa la media de la variable índice o latente (la variable original), $\frac{\partial E(y_i / x_i)}{\partial x_i} = \beta$.

Pero teniendo en cuenta la censura, el efecto marginal es

$$\frac{\partial E(y_i^* / x_i)}{\partial x_i} = \beta \Phi\left(\frac{x_i' \beta}{\sigma}\right) \quad (3.1.5)$$

Mc.Donald y Moffit (1980) propusieron una descomposición de $\partial E(y_i^* / x) / \partial x$ que resulta

$$\frac{\partial E(y_i^* / x_i)}{\partial x_i} = \beta [\Phi_i (1 - \lambda_i (\alpha_i + \lambda_i)) + \phi_i (\alpha_i + \lambda_i)], \quad (3.1.6)$$

siendo $\phi_i = \Phi(\beta' X_i / \sigma) = \Phi(\alpha_i)$ y $\lambda_i = \phi_i / \Phi_i$.

Considerando las dos partes por separado, se obtiene la siguiente descomposición de la pendiente

$$\frac{\partial E(y_i^* / X)}{\partial X} = P(y_i^* > 0) \frac{\partial E(y_i^* / X, y_i^* > 0)}{\partial X} + E(y_i^*, y_i^* > 0) \frac{\partial P(y_i^* > 0)}{\partial X}. \quad (3.1.7)$$

Así, un cambio en X produce dos efectos: afecta la media condicional de la respuesta en la parte positiva de la distribución y afecta la probabilidad de que la observación caiga en esa parte de la distribución.

Estimación por máxima verosimilitud

El modelo (3.1.3) a estimar es

$$y_i = \begin{cases} x_i' \beta + \epsilon_i & \text{si } y_i > 0 \\ 0 & \text{si } y_i \leq 0 \end{cases} \quad (3.1.8)$$

Sea S el conjunto de observaciones en los que se observa $y_i = 0$. Si el término de error es $N(0, \sigma^2)$ y denotando con ϕ y Φ sus funciones de densidad y de distribución, respectivamente, la función de verosimilitud es

$$L = \prod_S (1 - \Phi(x_i' \beta; \sigma^2)) \prod_{S^c} \phi(y_i - x_i' \beta; \sigma^2) \quad (3.1.9)$$

La estimación de los parámetros de los coeficientes de regresión por el método de máxima verosimilitud consiste en maximizar el logaritmo de la verosimilitud

$$\ln L = \sum_{y_i > 0} -\frac{1}{2} \left[\ln(2\pi) + \ln \sigma^2 + \frac{(y_i - x_i' \beta)^2}{\sigma^2} \right] + \sum_{y_i = 0} \ln \left[1 - \Phi\left(\frac{x_i' \beta}{\sigma}\right) \right]. \quad (3.1.10)$$

En esta expresión hay una parte similar a la que aparece en la regresión clásica, que corresponde a las observaciones no límite y otra que recoge las probabilidades de las observaciones límite. Éste es un tipo de verosimilitud no muy habitual, pues proviene de una mezcla de una distribución discreta y una continua. Los estimadores se obtienen derivando esa función con respecto a β y σ^2 .

Estimación por mínimos cuadrados

La esperanza condicional de la variable censurada y_i^* , cuya expresión es $E(y_i^* / x_i) = x_i' \beta + \sigma (\phi_i + (1 - \Phi_i) \lambda_i)$, es no lineal por lo que la estimación MCO de y_i^* incurre en un sesgo de variables omitidas.

Para solucionar el problema de la inconsistencia del estimador de MCO se propusieron varios estimadores consistentes alternativos al de máxima verosimilitud. Una posibilidad es aplicar mínimos cuadrados no lineales a la función de media condicional del modelo tobit utilizando toda la muestra, o bien a la función de media condicional de la regresión truncada utilizando sólo las observaciones no límite.

Otra posibilidad es utilizar la corrección en dos etapas del estimador de MCO propuesta por Heckman (1979). Una manera de obtener esta estimación a partir de la estimación por MCO es multiplicando a esta última por $1/P$, siendo P la proporción de observaciones no límite que hay en la muestra.

Los estimadores MCO son menores en valor absoluto que los máximo verosímiles. Pero corrigiendo el estimador MCO como se mencionó anteriormente, se obtiene una buena aproximación del estimador de máxima verosimilitud.

3.2. Prueba de exogeneidad

Esta prueba se utiliza para verificar si variables que se suponen exógenas en el sistema están correlacionadas con los residuos. Está basada en una modificación del contraste propuesto por Hausman para probar si en un modelo identificado y estimado existe algún problema de especificación.

Para estimar los parámetros de un modelo de regresión $Y = X\beta + u$ utilizando el método de MCO, se asume que las variables explicativas (X) son independientes de los errores (u). Si esto no se cumple, al menos una de las variables explicativas es endógena y el estimador no tiene las propiedades deseadas.

En el caso particular de someter a prueba una sola variable explicativa, la hipótesis a verificar es si X es exógena versus que es endógena. Para probarla se obtienen dos estimadores β_0 y β_1 que deben cumplir las siguientes condiciones:

β_0 es consistente y eficiente bajo H_0 pero no es consistente bajo H_1 ,

β_1 es consistente bajo H_0 y H_1 , pero no es eficiente bajo H_0 .

La estadística utilizada se construye a partir de $q = \beta_1 - \beta_0$ y $\text{Var}(q) = V_1 - V_0$ (donde $V_1 = \text{Var}(\beta_1)$ y $V_0 = \text{Var}(\beta_0)$), ambas calculadas bajo la hipótesis nula de exogeneidad), siendo σ_q^2 su estimador. Por lo tanto, bajo H_0

$$m = \frac{q^2}{\sigma_q^2} \text{ se distribuye asintóticamente como } \chi_1^2.$$

Por ejemplo, si $\beta_0 = \beta_{MCO}$, es decir, todas las X se consideran exógenas, y $\beta_1 = \beta_{VI}$, donde β_{VI} es el estimador MC2E considerando una de las X como endógena y las restantes X exógenas como variables instrumentales, entonces,

$$m = (\beta_{MCO} - \beta_{VI})' [\text{Var}(\beta_{VI}) - \text{Var}(\beta_{MCO})]^{-1} (\beta_{MCO} - \beta_{VI}). \quad (3.2.1)$$

En general, si se someten a prueba k variables explicativas, β_0 , β_1 y q son vectores de dimensión $k \times 1$, V_1 y V_0 son matrices de dimensión $k \times k$, siendo la estadística $m = q' [\hat{V}(q)]^{-1} q$ que se distribuye asintóticamente como χ_k^2 .

4. Definiciones operacionales

4.1. Definición de la muestra bajo estudio

La muestra en estudio está compuesta por 271 parejas conyugales (sin importar la legalidad de la unión) que viven en un mismo hogar, entrevistadas en la EPH en el aglomerado urbano correspondiente a los partidos del Conurbano Bonaerense en la primera onda de 1998 y en las tres ondas siguientes (debido a que se piensa efectuar en el futuro un estudio de panel). Esta muestra no incluye aquellas parejas donde el varón es inactivo y de 65 o más años de edad.

Con el objetivo de contar con una base de datos que contenga la información correspondiente a las variables de interés para la muestra de hogares, se programó una macro en SAS. La misma permite, de una manera automática, seleccionar exclusivamente los hogares de interés y generar un registro para cada pareja conyugal. Una vez seleccionados los hogares, se estudian posibles inconsistencias en algunas variables tales como edad, sexo y escolaridad. En base a estas variables también se identifican las parejas que se mudan durante el período de relevamiento de la encuesta.

Se discrimina a los individuos que componen cada pareja conyugal por su estado ocupacional, debido al importante rol que juega esta clasificación en la explicación del ingreso. Esta información se presenta en la tabla siguiente:

Tabla 1: Parejas según el estado ocupacional por sexo.

Estado ocupacional del varón	Estado ocupacional de la mujer			
	Ocupada	Desocupada	Inactiva	Total
Ocupado	87 (32,1%)	12 (4,4%)	137 (50,6%)	236 (87,1%)
Desocupado	11 (4,1%)	4 (1,5%)	8 (2,9%)	23 (8,5%)
Inactivo	2 (0,7%)	0 (0,0%)	10 (3,7%)	12 (4,4%)
Total	100 (36,9%)	16 (5,9%)	155 (57,2%)	271 (100,0%)

Los porcentajes de desocupación de la muestra en estudio con respecto a la población activa (8,9% para el varón y 13,8% para la mujer) son menores que las tasas de desocupación del aglomerado (14,9% varones, 17,7% mujeres). Esto se debe a que se estudia una muestra proveniente de una subpoblación (parejas conyugales) de la población total.

4.2. Definición de las variables

En este estudio se desea modelar el logaritmo del ingreso de la fuente laboral de cada uno de los cónyuges, en función de ciertas variables de interés.

Variables endógenas

- Y_1 : ingreso de fuente laboral del cónyuge (varón).
- Y_2 : ingreso de fuente laboral de la cónyuge (mujer).

También se trabaja con las siguientes variables transformadas:

- liflv : logaritmo del ingreso de fuente laboral del cónyuge (varón).

- **liflm**: logaritmo del ingreso de fuente laboral de la cónyuge (mujer).

En los casos en que los ingresos presentan valores iguales a cero, no se aplica la transformación y permanecen como tales. Cada una de estas variables presenta una distribución asimétrica con datos censurados a la izquierda, correspondientes a dichos ingresos cero. En el caso de Y_1 existe aproximadamente un 10% de datos censurados y en el caso de Y_2 , aproximadamente un 63%. Esto último indica un alto grado de no participación de las mujeres en el mercado laboral.

Si se analizan los ceros de acuerdo al estado ocupacional, se observa un comportamiento distinto entre los varones y las mujeres. La información se presenta en la Tabla 2.

Tabla 2: Distribución de los cónyuges con ingreso cero por sexo y estado ocupacional.

Estado ocupacional	Sexo	
	Varón	Mujer
Ocupado	2 (7,4%)	4 (2,3%)
Desocupado	13 (48,2%)	13 (7,6%)
Inactivo	12 (44,4%)	155 (90,1%)
Total	27 (100,0%)	172 (100,0%)

Se observa que la cantidad de mujeres con ingreso cero es mucho mayor que en el caso de los varones. Mientras que el ingreso cero en el hombre se debe principalmente al hecho de estar desocupado (48,2%), lo que se podría interpretar como ingreso cero "no voluntario", en la mujer se debe al hecho de estar inactiva (90,1%), lo que indicaría un ingreso cero "voluntario".

Variables exógenas

Las variables registradas en la EPH utilizadas para explicar el ingreso de fuente laboral (o su transformación logarítmica) denominadas variables exógenas (X), se definen para cada uno de los cónyuges. Para el caso de los varones se agrega a cada variable el sufijo **v** y para las mujeres se agrega el sufijo **m**.

- **estado ocupacional**: variable categórica que clasifica al encuestado como ocupado, desocupado o inactivo. Se trabaja con dos variables indicadoras ("dummy") las cuales resultan: **e1**=1 para desocupado, **e2**=1 para inactivo, siendo "ocupado" la categoría de referencia.
- **edad**: variable continua calculada a partir de la fecha de nacimiento declarada por el encuestado.
- **edad2**: variable edad elevada al cuadrado.
- **medesco**: promedio de los años de escolaridad declarados por el encuestado en las cuatro ondas.
- **lifnl**: logaritmo del ingreso de fuente no laboral. Variable continua que es el logaritmo de la suma de todos los ingresos no laborales declarados por el encuestado. Si es cero, permanece como tal.
- **jefe**: variable indicadora ("dummy") que toma el valor 1 para el cónyuge que se declara jefe del hogar en la encuesta.

Se cuenta con cada una de las siguientes variables exógenas relativas al hogar.

- **lotinghog**: logaritmo de otros ingresos del hogar. Variable continua que es el logaritmo de la suma de los ingresos laborales y no laborales de los restantes miembros del hogar. Si es cero, permanece como tal.
- **cont6**: cantidad de hijos menores de 6 años.
- **cont18**: cantidad de hijos entre 6 y 18 años.

4.3. Modelo planteado

Para modelar el ingreso laboral de la pareja conyugal se considera que cada ingreso individual es explicado de la siguiente manera:

$$\begin{cases} Y_1 = X_1 \beta_{11} + Y_2 \beta_{21} + v \\ Y_2 = X_2 \beta_{12} + Y_1 \beta_{22} + u \end{cases} \quad (4.3.1)$$

siendo:

Y_1 : vector de la variable endógena para el varón,

Y_2 : vector de la variable endógena para la mujer,

X_1 : matriz de características medibles (variables exógenas) que afectan a Y_1 ,

X_2 : matriz de características medibles (variables exógenas) que afectan a Y_2 ,

v y u : términos de error aleatorio distribuidos normalmente, que pueden estar correlacionados entre ellos.

5. Comparación de las estimaciones por distintos métodos

Con el fin de comparar los resultados de los distintos métodos se estiman las ecuaciones para los ingresos de fuente laboral del varón y la mujer que componen la pareja conyugal. Los métodos son:

5.1. Métodos uniecuacionales

Estiman cada ecuación por separado sin tener en cuenta la endogeneidad del ingreso del otro cónyuge.

5.1.1. Mínimos cuadrados ordinarios. Las estimaciones por este método no son consistentes por varias razones, entre las que se pueden destacar: a) la distribución de la variable respuesta no es normal; b) no se tiene en cuenta la endogeneidad del ingreso del otro cónyuge.

5.1.2. Regresión censurada. Este método tiene en cuenta el inciso a) del párrafo anterior, pero no tiene en cuenta el punto b). Se utiliza el procedimiento Tobit para estimar la ecuación que relaciona el ingreso con las demás variables de interés, para cada uno de los cónyuges por separado, debido a que la variable ingreso para ambos está censurada. Como se mencionó anteriormente, el porcentaje de censura es alto. La distribución que se utiliza para estimar esa relación es la lognormal.

5.2. Método de estimación en dos etapas

Otra forma de proceder con la estimación del ingreso de cada cónyuge por separado en el sistema de ecuaciones (4.3.1), consiste en una estimación en dos etapas. El método propuesto por Jacobsen y Rayack (1996) consiste en estimar, por ejemplo, el ingreso de la mujer utilizando todas las variables explicativas (X_1 y X_2) de las ecuaciones (4.3.1), excepto las endógenas. A partir de los valores estimados del ingreso para la mujer, a los que suele llamarse ingreso índice o latente (Greene, 1998), se estima por MCO la ecuación para el ingreso del varón (primera ecuación de (4.3.1)). Este procedimiento se repite para el otro sexo.

Las etapas se explican a continuación:

Primera etapa. La variable endógena (ingreso de uno de los cónyuges) censurada se estima mediante el siguiente modelo de regresión,

$$y_i^* = X\beta + v^* \quad i = 1, 2, \quad (5.2.1)$$

usando como variables explicativas todas las variables exógenas (indicadas con X). Las distribuciones de Y_1 e Y_2 son asimétricas y no toman valores negativos, por lo que se supone que las variables tienen distribución lognormal (Gráfico 1). La misma distribución poseen los ingresos estimados y_1^* e y_2^* (Gráfico 2), por lo tanto, $\log y_1^*$ y $\log y_2^*$ se distribuyen normalmente (Gráfico 3).

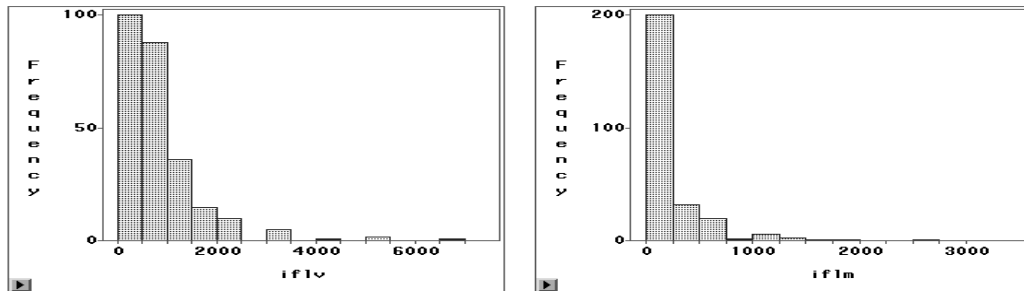


Gráfico 1: Distribución de los ingresos de fuente laboral del varón y de la mujer.

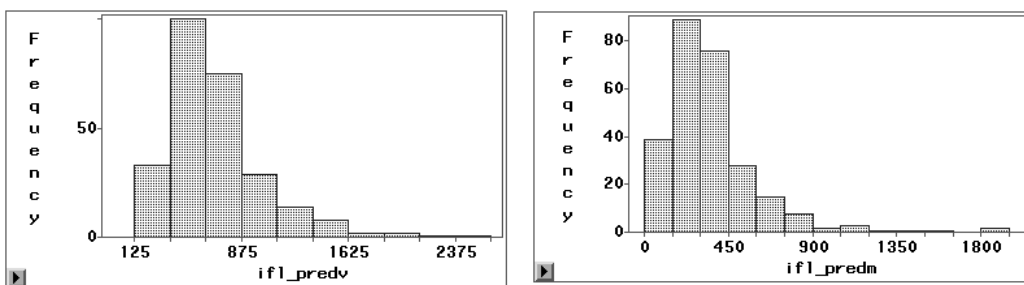


Gráfico 2: Distribución de los ingresos de fuente laboral del varón y la mujer estimados por regresión censurada.

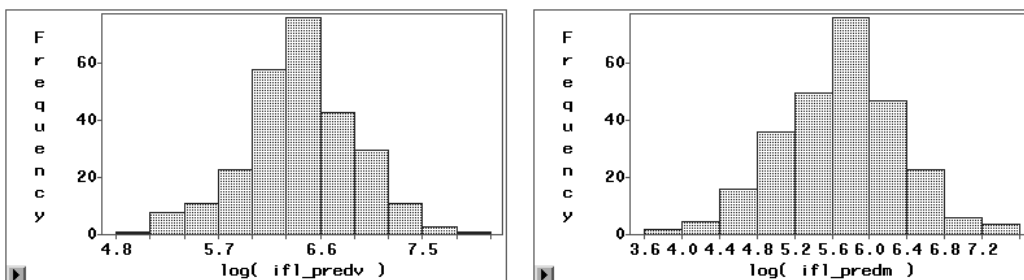


Gráfico 3: Distribución del logaritmo de los ingresos de fuente laboral del varón y la mujer estimados por regresión censurada.

Segunda etapa. Se estima por MCO la ecuación del logaritmo del ingreso para el otro cónyuge utilizando como variable explicativa el logaritmo de la estimada en la primera etapa.

$$\log y_j = \beta_{1j} X_j + \beta_{2j} \log \hat{y}_i^* + u_j \quad j = 1, 2, i = 1, 2, i \neq j. \quad (5.2.2)$$

Como todas las variables que intervienen para explicar y_j , también participan en y_i^* , la $\text{cov}(u, v^*) = 0$, se puede realizar sin problema la estimación por MCO de la segunda etapa.

En estas estimaciones existen al menos dos inconvenientes: a) como las variables Y_1 e Y_2 son censuradas, no se puede aplicar MCO en la segunda etapa; b) se estima cada ecuación de ingreso por separado sin tener en cuenta que conforman un sistema de ecuaciones.

5.3. Métodos multiecuacionales

El inconveniente planteado en el inciso b) del punto 5.2. es más difícil de resolver. Una solución sería encontrar la distribución condicional multivariada, equivalente a la distribución condicional univariada del modelo de Tobin, hallando sus respectivos momentos condicionales (tarea que se propone para el futuro).

Los métodos multiecuacionales que se utilizan son:

5.3.1. Mínimos cuadrados en tres etapas (MC3E). Sus estimaciones no son consistentes ya que las variables respuesta ($lifl$ y $liflv$) no cumplen el supuesto de normalidad, debido a la presencia importante de ceros.

5.3.2. Sistema "en dos etapas". Este método se indica con MC3EE. Se modelan los logaritmos de los ingresos estimados en 5.2. (latentes) utilizando un sistema de ecuaciones.

$$\begin{cases} \log \hat{Y}_1^* = \mathbf{X}_1 \beta_{11} + \log \hat{Y}_2^* \beta_{21} + v \\ \log \hat{Y}_2^* = \mathbf{X}_2 \beta_{12} + \log \hat{Y}_1^* \beta_{22} + u \end{cases} \quad (5.3.2.1)$$

$$u \sim N(0, \sigma_u^2), \quad v \sim N(0, \sigma_v^2).$$

Como tales logaritmos se distribuyen aproximadamente normal, se puede aplicar MC3E o cualquier método equivalente. La estimación de este sistema indicaría cómo varían los ingresos laborales latentes ante un cambio de las variables explicativas.

En este último caso se tienen en cuenta ambas ecuaciones en forma simultánea y las variables respuestas son normales, si bien no se usa como variable respuesta la variable original sino una estimación de la misma.

Se cree que esto puede constituir una aproximación al modelo de ingreso, donde los signos y magnitudes de los coeficientes responderían a lo sustentado por la teoría económica. Si bien no son exactamente la propensión marginal de contribución de cada variable, brindarían una idea de dicha proporción debido a que se cumplen los supuestos del método de estimación.

6. Resultados

Las estimaciones obtenidas a través de los distintos métodos detallados en la sección 5 se presentan en las tablas 3 y 4.

La prueba de exogeneidad de Hausman se realizó para cada una de las ecuaciones (la de la mujer y la del varón). Se utilizaron los estimadores $\hat{\beta}_{MCO}$ y $\hat{\beta}_{VI}$ (estimador MC2E considerando el $lifl$ del otro cónyuge como endógena). En ambos casos se rechazó H_0 y en consecuencia el ingreso del otro cónyuge debe considerarse como variable endógena.

De los resultados hallados se puede destacar que para el caso del varón:

- El coeficiente del logaritmo de otros ingresos del hogar ($lotinghog$) sólo resulta negativo y significativo cuando se lo estima por MC3EE.
- El coeficiente de mujer desocupada ($e1m$) es positivo en todos los modelos uniecuacionales mientras que resulta negativo cuando se usan modelos multiecuacionales, siendo además significativo para MC3EE.

Tabla 3: Estimación de los coeficientes por distintos métodos para el varón.

Variables	Modelos				
	Uniecuacionales			Multiecuacionales	
	MCO (4.1.1.)	Regresión Censurada (4.1.2.)	Tobit 2 etapas (4.2.)	MC3E (4.3.1.)	MC3EE (4.3.2.)
intercepto	1,388	1,485	4,626***	2,304	3,216***
e1v	-3,392***	-3,646***	-3,350***	-3,497***	-0,641***
e2v	-5,485***	-13,593***	-5,208***	-5,444***	-0,094***
edadv	0,065	0,048	0,031	0,060	0,060***
Edad2v	-0,001	-0,003	-0,000	-0,001	-0,001***
medescov	0,087***	0,084***	0,122***	0,086***	0,070***
lifnlv	-0,207***	-0,268***	-0,260***	-0,215***	-0,120***
jefev	0,755	0,840	1,299**	0,753	0,081***
lotinghog	0,009	0,012	0,009	0,008	-0,008***
Cont6	-0,025	-0,001	-0,181	-0,021	-0,020***
Cont18	0,087	0,117	0,148*	0,081	-0,004**
medescom	0,063**	0,072**	0,122***	0,666	0,014***
lifnlm	-0,143	-0,142	-0,177**	0,744	0,030***
Liflm	0,186***	0,192***	-0,549**	0,033	0,197***
e1m	1,332***	1,401***	0,324	-0,140	-0,165***
e2m	1,592***	1,631***	0,524***	0,067**	0,138***

* significativo al 10%, ** significativo al 5%, *** significativo al 1%.

Tabla 4: Estimación de los coeficientes por distintos métodos para la mujer.

Variables	Modelos				
	Uniecuacionales			Multiecuacionales	
	MCO (4.1.1.)	Regresión Censurada (4.1.2.)	Tobit 2 etapas (4.2.)	MC3E (4.3.1.)	MC3EE (4.3.2.)
intercepto	4,009***	3,720	25,711***	4,615***	14,422***
e1m	-4,445***	-5,465***	-5,519***	-4,357***	-0,883***
e2m	-5,658***	-19,490	-4,897***	-5,558***	0,333***
edadm	0,024	-0,009	0,267***	0,034	0,137***
edad2m	-0,000	0,000	-0,002***	-0,000	-0,001***
medescom	0,015	0,005	0,400***	0,026	0,197***
lifnlm	0,050	0,006	0,159*	0,032	-0,100***
jefem	0,115	0,584	-1,462**	-0,017	-1,875***
lotinghog	-0,013	-0,000	-0,070**	-0,011	-0,045***
cont6	0,084	0,341	-0,175	0,076	-0,303***
cont18	-0,082	-0,212	0,049	-0,076	0,183***
medescov	-0,018	-0,017	0,202***	-0,003	0,285***
lifnlv	-0,014	-0,000	-0,764***	-0,050	-0,463***
liflv	0,159***	0,261**	-4,924***	-0,012	-2,628***
e1v	-0,149	-0,374	-3,498***	-0,744	-1,161***
e2v	1,098**	2,170	0,322	0,180	0,345***

* significativo al 10%, ** significativo al 5%, *** significativo al 1%.

Para el caso de la mujer:

- El coeficiente de la mujer inactiva ($e2m$) estimado por MC3EE resulta ser positivo y significativo, mientras que por cualquier otro método es negativo.
- El coeficiente del logaritmo del ingreso de fuente no laboral de la mujer ($lifnlm$) es negativo y significativo sólo cuando se emplea MC3EE.
- Los únicos métodos con los que se obtiene un coeficiente negativo para las mujeres jefas de hogar ($jefem$) son Tobit dos etapas o los que tienen en cuenta la simultaneidad de los ingresos de los cónyuges.
- El único método que proporciona un coeficiente negativo y significativo para la cantidad de hijos menores de 6 años ($cont6$) y un coeficiente positivo y significativo para la cantidad de hijos entre 6 y 18 años ($cont18$) es MC3EE.
- El coeficiente positivo y significativo para la escolaridad del marido ($medescov$) sólo ocurre con los métodos Tobit dos etapas y MC3EE.
- El coeficiente del logaritmo del ingreso de fuente laboral del marido ($liflv$) es negativo y significativo cuando se utilizan Tobit dos etapas o MC3EE.

Los coeficientes estimados por MC3EE muestran los signos esperados para las variables de capital humano (edad y escolaridad). En cambio, en las variables de ingreso el signo es distinto: el signo de $iflv$ cuando explica el $iflm$ es negativo (a mayor ingreso del hombre, menor ingreso de la mujer), mientras que el de $iflm$ cuando explica el $iflv$ es positivo (a mayor ingreso de la mujer, mayor ingreso del varón). En cuanto a las variables ocupacionales de la mujer $e1m$ y $e2m$, se puede destacar que sus coeficientes tienen signos negativo y positivo respectivamente. Ello indicaría que las mujeres inactivas tienen ingresos potenciales superiores a los que podrían obtener en el mercado de trabajo.

7. Consideraciones Finales

En este trabajo introductorio se presentan varios métodos para la estimación de modelos que intentan explicar los ingresos de cada uno de los integrantes de la pareja conyugal. De acuerdo a los resultados obtenidos, es necesario hacer un llamado de atención a la hora de extraer conclusiones ya que los métodos de estimación corrientes pueden conducir a interpretaciones erróneas.

El problema se profundiza en el caso de las mujeres, especialmente por el alto porcentaje que no participa en el mercado laboral. Esto se refleja en la cantidad de coeficientes que cambian de signo y/o significación según el método utilizado y en que los métodos tradicionales no proporcionan la interpretación esperada bajo la teoría económica.

En futuros estudios, para considerar las características personales no observadas y minimizar los problemas de estimación derivados de la endogeneidad de algunas variables, se trabajará con datos de panel incorporando la información de las tres ondas adicionales en las que se presentan las parejas analizadas en este trabajo. También se incorporarán otros aglomerados para comparar comportamientos.



8. Referencias

- Becker, Gary (1973) "A Theory of Marriage:Part 1", *Journal of Political Economy*, Vol.81, N° 4.
- (1974): "A Theory of Marriage:Part 2", *Journal of Political Economy*, Vol.82, N° 2 part2.
- (1983): *El Capital Humano*, Alianza Universidad, Madrid.
- Behrman, Jere, Birdsall, Nancy and Deolalikar, Anil (1995): "Marriage Markets, Labor Markets, and Unobserved Human Capital: An Empirical Exploration for South-Central India", *Economic Development and Cultural Change*. 43:3. pp. 585-602.
- Berlinski, S. G. (2000): "On the estimation and test specification of wage equations using GMM", First version, *Nuffeld College, Oxford University*.
- Boulier, Bryan L. and Rosenzweig, Mark (1984): "Schooling, Search, and Spouse Selection: Testing Economic Theories of Marriage and Household Behavior", *Journal of Political Economy*, Vol.92m n° 4.
- Greene, W.H. (1998): *Análisis Económico*, Prentice Hall.
- Griliches, Z., (1977): "Estimating the Returns to Schooling: Some Econometric Problems", *Econometrica*, 45(1), pp 1-22.
- Heckman, J. (1979) "Sample selection bias as a specification error" *Econometrica*, 47, 153-161.
- Jacobsen, J. y Rayack, W. (1996): "Do Men Whose Wives Work Really Earn Less?", *American Economic Review*, Vol 86(2), May, pp 269-273.
- Lam, D. (1988) "Marriage Markets and Assortative Mating with Household Public Goods", Theoretical Results and Empirical Implications, *The Journal of Human Resources*, XXIII. 4, pp. 462-487.
- McDonald, J. and Moffitt, R. (1980) "The uses of Tobit analysis" *Review of Economic and Statistics*, 62,318-321
- Mincer, J. (1970): "The distribution of labor income: a survey; With special reference to human capital approach", *Journal of Economic Literature*, Vol. VIII, N°1.
- Tobin, J. (1958): "Estimation of Relationships for Limited Dependent Variables", *Econometrica*, 26, pp. 24-36.